Bayesian Networks and Causal Ecumenism

David Kinney

Forthcoming in Erkenntnis.[*]

Abstract

Proponents of various causal exclusion arguments claim that for any given event, there is often a unique level of granularity at which that event is caused. Against these causal exclusion arguments, causal ecumenists argue that the same event or phenomenon can be caused at multiple levels of granularity. This paper argues that the Bayesian network approach to representing the causal structure of target systems is consistent with causal ecumenism. Given the ubiquity of Bayesian networks as a tool for representing causal structure in both philosophy of science and science itself, this result speaks in favor of the ecumenical view, and against rival exclusionary accounts. Gebharter's (2017) argument that the Bayes nets formalism is consistent with causal exclusion is considered and rebutted.

1 Introduction

Suppose that I form the desire to move my arm, and subsequently move my arm. There are at least two ways of describing this process. On the one hand, a mental event (I form the desire) causes an action (I move my arm). On the other hand, a neural event (a physical process occurs in my brain that accounts for my forming the desire) causes me to move my arm. This multi-level picture is not unique to the relationship between psychology and neurobiology. One could tell a similar story with respect to the relationship between economics and psychology, biology and chemistry, chemistry and physics, and so on. Putative multi-level causal relations can cross disciplines: one could tell a coarse-grained story about El Niño weather patterns having a particular effect on the

[*]I am grateful to Sander Beckers, Jonathan Birch, Luc Bovens, Paul Daniell, Chris Dorst, Christopher Hitchcock, Christian List, Philip Pettit, Katie Steele, David Watson, an audience at the 2018 meeting of the APA Central Division in Chicago, and several anonymous reviewers for feedback on various versions of this paper.

GDP of Australia, or one could tell a fine-grained story about certain combinations of sea surface temperatures and zonal winds having the same impact. The thesis that the same event can be caused at varying levels of granularity is defended by Jackson and Pettit (1988, 1990a, 1990b, 1992, see also Pettit 2017). This thesis is also implicit in Davidson's (1970) philosophy of mind. Following Jackson and Pettit, I label this view *causal ecumenism*.

There are those that disagree with this ecumenical approach and mount in response *causal exclusion* arguments against the view described above. The most famous causal exclusion arguments are given by Kim (1989, 2000), who holds that all genuine causation occurs at the most ontologically fundamental level of description. On this view, all higher-level properties cited in causal relations must be epiphenomenal, since denying this epiphenomenalism would lead to the putatively false result that all physical events are causally over-determined. Thus, the existence of fine-grained causal relations results in the *exclusion* of more coarse-grained relations. I will not directly address the metaphysical aspects of Kim's argument here, but focus instead on a recent revival of his approach by Gebharter (2017). Gebharter argues that the Bayes nets approach to causal modeling, which is described in more detail below, is consistent with exclusion in all cases. His argument rests on the result that models that represent both higher- and lower-level properties of the same event fail to satisfy axioms of the Bayes nets formalism.[1]

I am sympathetic to Hitchcock's (2012) argument that since the soundness of any causal exclusion argument is sensitive to one's favored theory of causation, we should be hesitant to take any argument for or against causal exclusion to be generally decisive. Rather, it seems that the best that we can do is determine whether or not a causal exclusion argument is sound within the context of a particular theory of causation. In both philosophy of science and the sciences themselves, *Bayesian networks* (or "Bayes nets")—as developed by Pearl (2000) and Spirtes et al. (2000)—are a powerful formalism for representing causal structure. Bayes nets also entail a semantics for determining when one event causes another. This semantics makes essential use of counterfactual conditionals about hypothetical interventions on the causal structure of the system described by a given Bayes net. These interventional conditionals are at the core of Woodward's (2003) highly influential account of causation. In this paper, I argue that the Bayes nets formalism, and the

[1]Thus, I leave to one side versions of the causal exclusion argument in an interventionist setting that rely on *proportionality* constraints on causation, e.g. List and Menzies (2009) or Hoffmann-Kolss (2014).

interventionist semantics for causation that it entails, is in keeping with Jackson-and-Pettit-style causal ecumenism. My conclusions are similar to those put forward in Raatikainen (2010), Eronen (2012), Zhong (2014), Woodward (2015), and Polger et al. (2018). However, these authors only consider the causal exclusion problem insofar as it pertains to Woodward's interventionist semantics in deterministic cases; they do not frame their arguments in terms of the Bayes nets formalism, nor do they discuss probabilistic cases. Thus, my conclusions here have a wider scope than these earlier arguments.

Here is the plan for this paper. In Section 2, I introduce the Bayes nets approach to representing the causal structure of systems, and detail how the Bayes nets axioms imply an interventionist semantics for causation. In Section 3, I argue that in the Bayes nets context, the causal exclusion problem is a problem of variable choice, and that the Bayes nets formalism supports a generally ecumenical response to causal exclusion arguments. In Section 4, I respond to Gebharter's argument that the Bayes nets formalism entails causal exclusion. In Section 5, I respond to the argument that ecumenism is false in a Bayes nets context because macro-level interventions on variables are not possible or well-defined. In Section 6, I offer concluding remarks.

2 Background

2.1 Bayesian Networks

A Bayes net is a triple $\mathcal{N} = \langle \mathcal{V}, \mathcal{E}, p(\cdot) \rangle$. $\mathcal{V}$ is a set of random variables that are all defined over the same event space, and whose values denote different possible states of the system being represented. $\mathcal{E}$ is an acyclic set of ordered pairs of the variables in $\mathcal{V}$. These ordered pairs, called *edges*, are usually represented visually as arrows pointing from one variable to another. The pair $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is called a *graph*. If a graph contains an edge $V_1 \rightarrow V_2$, then V_1 is a *parent* of V_2, and V_2 is a *child* of V_1. A chain of parent-child relationships is called a *directed path*. If there is a directed path from a variable V_3 to a variable V_4, then V_4 is a *descendant* of V_3 and V_3 is an *ancestor* of V_4. Finally, $p(\cdot)$ is a probability distribution defined over the Cartesian product of the ranges of each variable in $\mathcal{V}$. The edges of a Bayes net are meant to represent a relation of direct causal relevance between the variables that are related by an edge. To illustrate, suppose that a Bayes net contains an edge $X \rightarrow Y$. This is meant to represent the fact that if $X = x_i$ and $Y = y_j$, then the event represented

by $X = x_i$ is directly causally relevant to the event represented by $Y = y_j$.

This discussion raises a seemingly pertinent question: what exactly is meant by a relation of direct causal relevance? The theory of Bayes nets does not answer this question by providing a set of necessary and sufficient conditions for one event to be a direct cause of another. Instead, the use of Bayes nets to represent causal structure depends on a "non-reductive" or "Euclidean" approach to clarifying the nature of causation (see Hausmann and Woodward 1999; Spirtes et al. 2000, p. 3; Glymour 2004). Rather than providing necessary and sufficient conditions for X to be directly causally relevant to Y, if the graph in a Bayes net truthfully represents the causal relationships in a system, then we should expect the Bayes net to satisfy the following two axioms:

> **Causal Markov Condition (CMC):** A Bayes net $\mathcal{N} = \langle \mathcal{V}, \mathcal{E}, p(\cdot) \rangle$ satisfies CMC if and only if, according to $p(\cdot)$, each variable in $\mathcal{V}$ is probabilistically independent of its non-descendants in $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, conditional on its parents.
>
> **Minimality:** A Bayes net $\mathcal{N} = \langle \mathcal{V}, \mathcal{E}, p(\cdot) \rangle$ satisfies Minimality if and only if it satisfies CMC, and any Bayes net $\mathcal{N}^* = \langle \mathcal{V}, \mathcal{E}^*, p(\cdot) \rangle$ such that $\mathcal{E}^* \subset \mathcal{E}$ does not satisfy CMC.

The expectation that a Bayes net whose graph represents the true causal structure of a system will satisfy CMC is perhaps best justified by Pearl (2000, p. 30), who shows that if the value taken by any variable in a Bayes net is a function of the values taken by its parents and an error term that is not correlated with the error in any other variables in the graph, and the graph contains all common causes of two or more variables, then the probability distribution over the variables in the Bayes net will satisfy CMC. Thus, under the plausible assumption that events are determined by their direct causes, plus some exogenous error probability, we should expect a representation of the causal structure of a system to satisfy CMC.

The Minimality axiom is needed because CMC does nothing to restrict the number of edges that can be included in a graph. To illustrate, consider Figure 1. Let us assume that Yellow Fingers and Lung Cancer are independent conditional on Smoking. The inclusion of an edge between Yellow Fingers and Lung Cancer, in spite of the fact that they are independent conditional on Smoking, does not induce a violation of CMC. So CMC alone does not rule out a Bayes net that represents a scenario in which having yellow fingers causes lung cancer, even when the correlation between these two variables is entirely explained by the increased presence of smoking in those with both

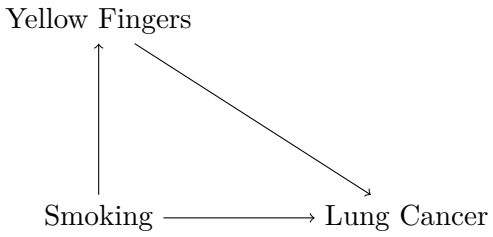

Figure 1: Graph with Excess Edge

yellow fingers and lung cancer. Thus, we need a further constraint on Bayes nets to ensure that they do not contain excess edges, i.e. an Ockham's razor condition to ensure that Bayes nets have the simplest structure needed to account for the observed data. The graph in Figure 1 violates Minimality, since removing the arrow between Yellow Fingers and Lung Cancer does not result in a violation of CMC, on the supposition that Yellow Fingers and Lung Cancer are independent, conditional on Smoking. By contrast, removing the arrow from Smoking to Lung Cancer would lead to a violation of CMC, on the supposition that Lung Cancer is not independent of Smoking, conditional on Yellow Fingers. In general, Minimality ensures that the edges in a Bayes net are only doing the work required in order to satisfy CMC. This axiom will be crucial to my response to Gebharter's exclusion argument in Section 4.[2]

In what follows, I will stipulate that certain Bayes nets are *true representations* of a given target system. This notion of a true representation can be made precise as follows. Let $\mathcal{V}_\Phi$ be a variable set such that the values of each variable $X \in \mathcal{V}_\Phi$ represent a different state or set of states of some system Φ. Let $p_\Phi(\cdot)$ be the probability distribution over the values of those variables that has maximum likelihood, given the observed data. Let us suppose further that the variable set $\mathcal{V}_\Phi$ is sufficient to represent the causal structure of the system in question; changes in the world that are not represented in the model and are common causes of two or more variables in the model do not make a difference to the probability distribution over $\mathcal{V}_\Phi$. A Bayes net $\mathcal{N}_\Phi = \langle \mathcal{V}_\Phi, \mathcal{E}_\Phi, p_\Phi(\cdot) \rangle$ is a true representation of Φ just in case it satisfies CMC and Minimality and its graphical structure matches the true causal structure of Φ, when the possible states of Φ are represented by a particular variable set. Determining the true causal structure of some system is difficult. Given a variable set and

[2]Many readers will be familiar with the Faithfulness condition for Bayes nets, which is strictly stronger than Minimality. I choose Minimality over Faithfulness as an adequacy condition for the causal interpretation of Bayes nets, since there is a case to be made that Bayes nets that satisfy Minimality but not Faithfulness are accurate representations of some causal systems. For a perspicuous comparison of the two conditions, see Zhang (2012).

a probability distribution over the values of those variables, standard causal modeling algorithms, such as those developed by Spirtes et al. (2000), often return more than one Bayes net that satisfies CMC and Minimality, with edges potentially pointing in different directions. Experiments can be helpful in ruling out certain graphical structures, so as to determine a single Bayes net that is an adequate representation of a target system. Additionally, sophisticated causal modeling techniques developed by Shimizu et al. (2006) and Janzing and Scholkopf (2010) may, in some cases, allow for the inference of a single representationally adequate causal graph from purely observational data. Going forward, I will make the idealizing assumption that for a given target system Φ, scientists are capable of determining, for a given variable set $\mathcal{V}_\Phi$ and joint probability distribution $p_\Phi(\cdot)$, a single causal graph that is a true representation of Φ.

Note that for a given target system, there can be multiple Bayes nets that truly represent the system, depending on the set of variables chosen. This reflects the fact that the set of values of each variable can be interpreted as a partition on some specification of the set of possible worlds Ω. For any such Ω, it is possible to have two sets of partitions/variables $\mathcal{V}_1$ and $\mathcal{V}_2$ that are non-equivalent, but are such that the true probability distribution will be Markov and Minimal to both of them.[3] To ask which of these is the single, true causal graph is to make a category mistake about the nature of causation as it is described by Bayes nets; recall above that the Bayes nets approach does not provide necessary and sufficient conditions for a causal relationship to exist, *simpliciter*. Rather, it counts as genuine all causal relationships that can be represented by a Bayes net that satisfies the Markov and Minimality conditions described above.

2.2 Intervention and Causation

The Bayes nets approach to causal modeling entails a semantics of causation in terms of interventional counterfactual conditionals. An *intervention* on a Bayes net is a (possibly hypothetical) exogenous change in the value(s) of a variable or group of variables in that Bayes net. By an 'exogenous change', I mean a change that occurs independently of the values taken by any other variables in the Bayes net, as though the change were achieved by a manipulation of the target system by factors external to the system.

Woodward (2003, p. 77) formalizes the connection between intervention and actual causation

[3]See Spirtes (2007) and Eberhardt (2016) for rigorous demonstrations of this point.

as follows. Let it be the case that x_i and y_j are possible values taken by variables X and Y in a Bayes net $\mathcal{N}$. For $X = x_i$ to be an actual cause of $Y = y_j$ in $\mathcal{N}$, the following conditions must hold:

(i) It is in fact the case that $X = x_i$ and that $Y = y_j$.

(ii) According to $\mathcal{N}$, Y is a descendant of X.

(iii) There is at least one directed path from X to Y such that, holding all parents of Y not on this path fixed at their actual values, there is some intervention setting X to x_k, where $x_i \neq x_k$, such that $Y = y_l$, where $y_j \neq y_l$.

Woodward's condition (iii) assumes a deterministic relationship between variables in a causal graph. However, this condition can be straightforwardly adopted into the probabilistic context as follows:

(iii*) According to the joint distribution $p(\cdot)$ over the variables in the graph $\mathcal{G}$, there is at least one directed path from X to Y such that, holding all parents of Y not on this path fixed at their actual values, $p(y_l|do(x_k)) > p(y_l)$, where $x_i \neq x_k$ and $y_j \neq y_l$.

The $do(\cdot)$ operator denotes that the variable X is set to the value x_k via an intervention. It might be argued at this point that (iii*) does not give us everything that what we want out of a probabilistic definition of actual causation, since we also want it to be the case that the cause $X = x_i$ raises the probability of its effect $Y = y_j$. It turns out that we can include this condition in our definition of actual causation, while also preserving the inclusion of (iii*) in such a definition, by replacing (iii*) with the following:

(iii**) According to the joint distribution $p(\cdot)$ over the variables in the graph $\mathcal{G}$, there is at least one directed path from X to Y such that, holding all parents of Y not on this path fixed at their actual values, $p(y_j|do(x_i)) > p(y_j)$.[4]

We can replace (iii*) with (iii**) while guaranteeing the truth of (iii*) because the following proposition is true:

[4]Note that, since we assume that the value of X is x_i, the interventional conditional probability $p(y_j|do(x_i))$ denotes the probability that $Y = y_j$ on the supposition that X has been set to x_i via an intervention, as opposed to the probability that $Y = y_j$ conditional on the observation that $X = x_i$.

Proposition 1. *If there is a Bayes net $\mathcal{N}$ with variables X and Y that take the values x_i and y_j and at least one directed path from X to Y such that, holding all parents of Y not on this path fixed at their actual values, $p(y_j|do(x_i)) > p(y_j)$, then there is some pair of values (x_k, y_l) such that, holding all parents of Y not on the same directed path fixed at their actual values, $p(y_l|do(x_k)) > p(y_l)$, where $x_i \neq x_k$ and $y_j \neq y_l$.*

Thus, (iii**) implies (iii*), and we can adopt Woodward's definition of actual causation into a probabilistic, Bayes nets context by adopting (i), (ii), and (iii**) as a definition of probabilistic actual causation.

This definition of probabilistic actual causation in terms of conditions (i), (ii), and (iii**) is subject to possible counterexamples. According to these counterexamples, intuitive judgements about what counts as the actual cause of some event is not the actual cause picked out by the definition. In particular, Fenton-Glynn (2017) provides several cases in which this naive definition of actual causation fails. However, my goal in this paper is not to provide an all-encompassing definition of actual causation. Indeed, Glymour et al. (2010) provide a compelling argument that such an account will not be produced, due to a combinatorial explosion of possible counterexamples to our intuitive judgements about causation that emerge as we add more variables to a Bayes net. Rather, I intend only to argue that conditions (i), (ii), and (iii**) succeed in identifying actual causal relations between the values taken by variables in a Bayes net in a range of scientifically realistic cases, and then show that, on such an account, causal ecumenism holds.

The definition of actual causation in terms of conditions (i), (ii), and (iii**) implies that in order to determine whether one event causes another in a probabilistic context, we need to be able to determine whether an intervention bringing about the first event increases the probability of the second event. The Bayes nets formalism provides an explicit methodology for determining whether these conditions for causation are in fact satisfied. To illustrate, let us begin by defining the *Markov factorization* of a Bayes net. Pearl (2000, p. 15-16) proves that if $\mathcal{V} = \{V_1, V_2, ..., V_m\}$, and if each variable in $\mathcal{V}$ has a corresponding value $v_1, v_2, ..., v_m$, and if $\mathbf{pa_{V_i}}$ is the vector of values taken by the set of parents of a variable V_i in the Bayes net $\mathcal{N} = \langle \mathcal{V}, \mathcal{E}, p(\cdot) \rangle$, and if $\mathcal{N}$ satisfies CMC, then

the probability $p(v_1, v_2, ..., v_m)$ can be factorized as follows.

$$p(v_1, v_2, ..., v_m) = \prod_{i=1}^{m} p(v_i|\mathbf{pa_{V_i}}) \tag{1}$$

Next, suppose that we intervene on a set of variables $\mathcal{X} \subseteq \mathcal{V}$, setting it to the vector of values $\mathbf{x_o}$. Pearl (2000, p. 30) and Spirtes et al. (2000, p. 51) show that in a Bayes net that satisfies CMC, the interventional conditional probability $p(v_1, v_2, ..., v_m|do(\mathbf{x_o}))$ can be obtained using the following truncated factorization:

$$p(v_1, v_2, ..., v_m|do(\mathbf{x_o})) = \prod_{i=1}^{m} p_{do(\mathbf{x_o})}(v_i|\mathbf{pa_{V_i}}) \tag{2}$$

Where each probability $p_{do(\mathbf{x_o})}(v_i|\mathbf{pa_{V_i}})$ is defined as follows:

$$p_{do(\mathbf{x_o})}(v_i|\mathbf{pa_{V_i}}) = \begin{cases} p(v_i|\mathbf{pa_{V_i}}) & \text{if } V_i \notin \mathcal{X} \\ 1 & \text{if } V_i \in \mathcal{X} \text{ and } v_i \text{ consistent with } \mathbf{x_o} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Huang and Valtorta (2006) show that this procedure can be used to calculate the probability distribution over any combination of variable values in a Bayes net that satisfies CMC, given any intervention. Thus, if a Bayes net satisfies CMC, we can use interventions to identify causal relationships, in keeping with Woodward's account of causation given above.

A less formal account of the connection between CMC and interventional conditional probability distributions can be stated as follows. Let $\mathcal{N} = \langle \mathcal{V}, \mathcal{E}, p(\cdot) \rangle$ be a Bayes net. If we intervene on some set of variables $\mathcal{X} \subseteq \mathcal{V}$, then we make it the case that the values of the variables in $\mathcal{X}$ no longer depend on their parents, but instead depend solely on the intervention. This can be represented graphically by a sub-graph of $\mathcal{N}$ in which all arrows into $\mathcal{X}$ are removed. This sub-graph is called the *pruned sub-graph* of $\mathcal{N}$ for an intervention on $\mathcal{X}$. Spirtes et al. (2000) prove that this pruned sub-graph of $\mathcal{N}$ satisfies CMC with respect to the intervention distribution $p_{do(\mathbf{x_0})}(\cdot)$, so that we can calculate the joint probability distribution over the pruned sub-graph of $\mathcal{N}$ given an intervention on any set of variables $\mathcal{X}$, using the equation (2). This calculation allows us to determine which types

of events represented in a Bayes net cause other types of events represented in the same Bayes net, since we can use interventions both to change the values of variables and to hold selected variables fixed at their actual values.

This completes my exegesis of the Bayes nets approach to causal representation, and how it facilitates an interventionist semantics of causation. In what follows, I show that this approach is consistent with an ecumenist response to various causal exclusion arguments.

3 The Case for Ecumenism in Bayes Nets

Causal ecumenism is the claim that the same event can be caused at multiple levels of granularity. This claim can be made more precise, in order to be substantively defended. Let X be any random variable. Let X be a *coarsening* of some other random variable X' just in case X is not identical to X', and the range of X is a *quotient set* of the range of X'. One set A is a quotient set of another set B just in case A contains all and only the equivalence classes of B according to some equivalence relation defined over B. Thus, if one variable is a coarsening of another, then the more coarse-grained variable *supervenes* on the more fine-grained variable; there can be no change in the value of the coarse-grained variable without a subsequent change in the value of the fine-grained variable. Throughout this paper, if one variable is a coarsening of another, then the more fine-grained variable will be indicated via a superscript $\prime$. If X is a coarsening of X', then X' is a *refinement* of X. We can now state the thesis of causal ecumenism in the Bayes nets context as follows.

> **CE:** Let it be the case that $C' = c'_l$ causes $E' = e'_s$ according to a Bayes net $\mathcal{N}' = \langle \mathcal{V}', \mathcal{E}', p(\cdot) \rangle$, where $\mathcal{N}'$ is a true representation of some target system Φ, and where $\mathcal{N}'$ satisfies CMC and Minimality, and where $E' \in \mathcal{V}'$ and $C' \in \mathcal{V}'$. In some cases in which these preliminary conditions are satisfied, there is a Bayes net $\mathcal{N} = \langle \mathcal{V}, \mathcal{E}, p(\cdot) \rangle$ that is also a true representation of Φ such that: 1) $\mathcal{V} = \{C\} \cup \mathcal{V}' \setminus \{C'\}$, 2) the joint probability distribution $p(\cdot)$ is defined over all variables in $\mathcal{V} \cup \{C'\}$, 3) $\mathcal{N}$ satisfies CMC and Minimality, 4) C is a coarsening of C', 5) if $C' = c'_l$ implies that $C = c_j$, then $C = c_j$ causes $E' = e'_s$.

Thus, under causal ecumenism, if we can identify a cause of some event at one level of granularity,

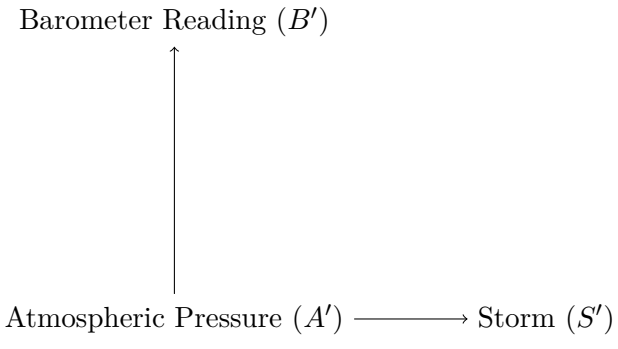

Figure 2: Weather System Graph with Fine-Grained Pressure Variable

$p(A' = low, B' = high, S' = yes) = .042$
$p(A' = low, B' = high, S' = no) = .003$
$p(A' = low, B' = low, S' = yes) = .267$
$p(A' = low, B' = low, S' = no) = .021$
$p(A' = medium, B' = high, S' = yes) = .1$
$p(A' = medium, B' = high, S' = no) = .1$
$p(A' = medium, B' = low, S' = yes) = .067$
$p(A' = medium, B' = low, S' = no) = .067$
$p(A' = high, B' = high, S' = yes) = .042$
$p(A' = high, B' = high, S' = no) = .267$
$p(A' = high, B' = low, S' = yes) = .003$
$p(A' = high, B' = low, S' = no) = .021$

Table 1: Joint Distribution for Fine-Grained Weather System

then it is sometimes possible that we can identify a cause of the same event at a coarser level of granularity, by replacing the causal variable with a coarsening of itself.

As causal ecumenism is an existentially quantified thesis, its truth can be demonstrated by an example. Suppose that the graph in Figure 2 is an adequate representation of the causal structure of the weather system in a given region. A' is a variable representing atmospheric pressure, with possible values $\{low, medium, high\}$. B' is a variable representing the reading of a barometer, with possible values $\{low, high\}$. S' is a variable representing whether or not a storm occurs, with possible values $\{yes, no\}$. The joint probability distribution over these variables is given in Table 1. Examining this joint distribution, one can see that low pressures, low barometer readings, and the presence of storms are correlated, as are high pressures, high barometer readings, and the absence of storms. However, there is some measurement error, such that there is not a perfectly deterministic relationship between atmospheric pressure and readings. Similarly, there is some stochasticity in the relationship between atmospheric pressure and the presence or absence of storms. One can check that this distribution satisfies CMC with respect to the graph shown in Figure 2. Thus, we are able to calculate the conditional probability distribution over variables in the graph, given interventions on other variables.

Suppose that $A' = low$ and $S' = yes$. The following calculation shows that low atmospheric pressure causes storms (details of this calculation, and all subsequent calculations, are given in the appendix):

$$p(S' = yes|do(A' = low)) = .93 > p(S' = yes) = .52 \tag{4}$$

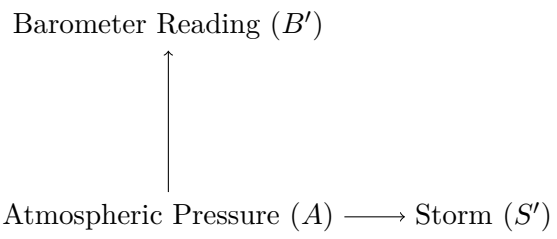

Figure 3: Weather System Graph with Coarse-Grained Pressure Variable

$p(A = low/medium, B' = high, S' = yes) = .142$
$p(A = low/medium, B' = high, S' = no) = .103$
$p(A = low/medium, B' = low, S' = yes) = .334$
$p(A = low/medium, B' = low, S' = no) = .088$
$p(A = high, B' = high, S' = yes) = .042$
$p(A = high, B' = high, S' = no) = .267$
$p(A = high, B' = low, S' = yes) = .003$
$p(A = high, B' = low, S' = no) = .021$

Table 2: Joint Distribution for Coarse-Grained Weather System

By contrast, the following calculation shows that a low barometer reading does not cause storms:

$$p(S' = yes|do(B' = low)) = .52 = p(S' = yes) = .52 \tag{5}$$

Thus, this case is consonant with a commonsense understanding of meteorology; while low atmospheric pressure can cause a storm to occur, a low reading on a barometer cannot bring about the same effect.

Now suppose that we coarsen the variable A'. To do this, we define an equivalence relation over the range $\{low, medium, high\}$ such that 'low' and 'medium' are equivalent to each other as well as themselves, and 'high' is equivalent only to itself. Defining a quotient set of the fine-grained range according to this equivalence relation results in a coarse-grained variable A with the range $\{low/medium, high\}$. If we replace the variable A' in Figure 2 with the variable A, we get the graph in Figure 3 and the joint distribution over that graph shown in Table 2. This joint distribution assumes that $p(A' = low|do(A = low/medium)) = p(A' = medium|do(A = low/medium))$; this assumption is made for mathematical tractability and is not generally required for any subsequent results. This joint distribution is calculated under the stipulation that if $\{x'_1, x'_2, ..., x'_k\}$ are all and only those values of a fine-grained variable that stand in some equivalence relation to each other and are all mapped to the coarse-grained variable value x_r, then $p(x_r) = \sum_{i=1}^{k} p(x'_i)$. One can check that the joint distribution in Table 2 satisfies CMC with respect to the graph in Figure 3. Thus we can confirm that if $A = low/medium$, then this is a cause of $S' = yes$, by calculating the following:

$$p(S' = yes|do(A = low/medium)) = .714 > p(S' = yes) = .52 \tag{6}$$

This calculation shows that causal ecumenism is satisfied in this case. The event represented by $A' = low$ causes the event represented by $S' = yes$ to occur, according to the Bayes net formed by the graph in Figure 2 and the joint distribution in Table 1. Further, the Bayes net formed by the graph in Figure 3 and the joint distribution in Table 2 is such that: 1) the only difference in variable sets between the two graphs is the replacement of A' with its coarsening A, 2) the joint probability distribution $p(\cdot)$ is defined over A, A', S' and B', 3) the Bayes net satisfies CMC and Minimality, 4) A is a coarsening of S', and 5) $A' = low$ implies that $A = low/medium$ and $A = low/medium$ causes $S' = yes$. Thus, all the conditions for causal ecumenism are satisfied.

At this point, one might wish to raise the objection of why we do not consider a graph that includes both A and A', where these two variables may be related by a common cause, or directly causally related in either direction. I note for now that I will return to this objection in Section 4, in my response to Gebharter's causal exclusion argument.

There is a more general result that speaks in favor of causal ecumenism. Specifically, the following proposition is true:

Proposition 2. *For any Bayes net $\mathcal{N}' = \langle \mathcal{V}', \mathcal{E}', p(\cdot) \rangle$ that satisfies CMC and Minimality, and any variable $C' \in \mathcal{V}'$ such that C' has more than two values in its range, if $C' = c'_l$ causes $E' = e'_s$, where $E' \in \mathcal{V}'$, then there is a Bayes net $\mathcal{N} = \langle \mathcal{V}, \mathcal{E}, p(\cdot) \rangle$ that satisfies the five conditions listed in **CE**.*

This proposition shows that any variable with more than two values can be coarsened into a variable with fewer values, while preserving any causal relations between the value taken by the coarsened variable and the value taken by other variables in the graph. If, in this coarsening process, the value c'_l is deemed to be equivalent to some other fine-grained value c'_u, then replacing $C' = c'_l$ with the more coarse-grained cause $C = c_j$, where $c_j = \{c'_l, c'_u\}$, can result in a more "abstract" causal representation than its fine-grained counterpart, in Weslake's (2010) sense of having more possible instantiations in various target systems that the model might be used to represent. This occurs in the case presented above; $A' = low$ and $A' = medium$ were deemed equivalent, and the causal relation that resulted from grouping them together in a coarsening ($A = medium/low$ causes $S' = yes$) was more abstract. Additionally, the existence of a more coarse-grained version of some fine-grained causal claim implies that the reverse is also true; for some coarse-grained variables, one

can find a way of refining the values of that variable so as to derive a more fine-grained causal claim (providing that some hypothetical limit to the fine-grainedness of the description of any system has not been reached).

To clarify, Proposition 2 does not imply that any causal claim can be restated in more coarse-grained, abstract language. Rather, it says only that variables with more than two values can be replaced by some coarsening while maintaining our ability to derive accurate causal claims from a Bayes net model of the target system. To illustrate this distinction, suppose that in some Bayes net representing whether or not flooding will occur in a region, there is a three-valued variable R' representing the amount of rainfall in a region, with possible values $\{low, medium, high\}$, and a two-valued variable F' representing whether a flood occurs, with possible values $\{no, yes\}$. Suppose further that $R' = high$ causes $F' = yes$. Under some joint probability distributions, it may be the case that if we coarsen R' into R so that it can take the values $\{low, medium/high\}$, then it is not the case that $R = medium/high$ causes $F' = yes$. Similarly, if we coarsen R' into a variable R that can take the values $\{low/high, medium\}$, then it may not be the case that $R = low/high$ causes $F' = yes$. Having ruled out these options, Proposition 2 does imply that we can coarsen R' into a variable R that can take the values $\{low/medium, high\}$, such that $R = high$ causes $F' = yes$. This leaves us with effectively the same causal claim as we had pre-coarsening, albeit one that is embedded within a more coarse-grained model of the target system as a whole.

Additionally, Proposition 2 does not imply that *any* coarsening of the causal variable will preserve the causal relationships between values taken by the causal variable and the values taken by its effect variable(s). Indeed, some poorly-chosen coarsenings will result in a situation such that interventions on a fine-grained variable X' change the probability distribution over another variable Y', but when X' is replaced by a coarsening X, no interventions on X change the probability distribution over Y'. Consider the following case, based on an example due to Spirtes and Scheines (2004). In humans, the presence of high-density lipids in the bloodstream decreases the risk of heart attack, whereas the presence of low-density lipids in the bloodstream increases the risk of heart attack. Let L' be a variable whose range contains pairs of real numbers denoting the amount of each lipid (in mg/dL) in a patient's bloodstream. So if a patient's blood contains 100 mg/dL of low-density lipids and 160 mg/dL of high-density lipids, then $L' = \langle 100, 160 \rangle$. Let H' be a variable representing whether or not someone has a heart attack. Clearly, there are some interventions on L'

such that conditioning on those interventions increases the probability that $H' = yes$ or $H' = no$. However, suppose that we replace L' with its coarsening L, where the value of L is just the sum of the patient's two lipid levels (e.g. if $L' = \langle 100, 160 \rangle$, then $L = 260$). There are marginal probability distributions over the fine-grained variable L' such that, for any intervention on L, conditioning on that intervention does not raise or lower the probability of heart attack. This is because the total amount of lipids in the patient's bloodstream is not informative as to the patient's risk of heart attack if we do not also know the density of those lipids. These sorts of cases show that in choosing coarse-grained variables, scientists have to be careful to choose coarsenings that still allow us to derive causal claims in the more coarse-grained model. Proposition 2 shows that this can be done in principle, as long as the coarsening is defined in terms of an appropriate equivalence relation between values of the fine-grained variable.

However, this example does bring to the foreground an important feature of causal explanation and causal ecumenism. Suppose that, in general, hypothetical interventions that set a person's overall lipid count to high levels tend to do so by increasing the amount of low-density lipids in that person's bloodstream, while leaving fixed the amount of high-density lipids. Thus, a hypothetical intervention setting L to a high value, e.g. $L = 300$, would be evidence of a person being more likely to have a high level of low-density lipids. This would make it the case that $p(H' = yes|do(L = 300)) > p(H' = yes)$, and the unconditional joint probability distribution over the Bayes net would have to be such that this interventional conditional probability can be derived from it. If we assume that L and H' are variables in a Bayes net that satisfies CMC and Minimality, then we get the result that if someone has a heart attack and their total lipid level is 300 mg/dL, then their total lipid level caused them to have a heart attack.

At first glance, this seems like a strange result. After all, interventions on the total lipid count only raise a person's likelihood of having a heart attack in virtue of the fact that these interventions tend to also increase the amount of low-density lipids in their bloodstream. So really, it seems that it is this fact about low-density lipids, rather than the bare fact of the person's overall lipid count, that explains their heart attack. However, according to the ecumenism that I have defended above, the correct analysis of this case is as follows. It is true that, on a more fine-grained model, it is really the person's high level of low-density lipids that causes their heart attack. However, it is also the case that, on a more coarse-grained model, their total lipid count causes their heart attack. To

say that both causal claims can be true is just to accept the thesis of causal ecumenism. However, this does not entail that the two models furnish *equally good explanations* of the person's heart attack. Indeed, I would hold that in this case, the more fine-grained explanation is the better one. It is just that both putative explanations do satisfy the conditions to be an explanation of the person's heart attack.

Biting this bullet in these cases allows us to give the same response in cases of scientific explanation at different levels of granularity where such a response appears to be warranted. Suppose that we heat a box of gas, and thereby cause an increase in the pressure that the gas exerts on the box. We can explain the increased pressure on the box by citing the increase in heat brought about via an intervention. At a more fine-grained level, we can say that the increased average kinetic energy of the particles of gas in the box explains the pressure that the gas exerts on the box. However, both of these explanations only make sense under the assumption that the increase in average kinetic energy of the particles in the box is at least somewhat evenly distributed between the particles. It may be that, as a result of heating, the particles in the center of the box are very high-energy, and the particles further from the center of the box maintain their kinetic energy from before the box is heated. Average kinetic energy increases, but is not evenly distributed. Under these conditions, the pressure on the box would not increase, since particles near the edge of the box would not be colliding into it with greater velocity.

It is scientifically respectable to explain the increase in the pressure on the interior of the box by citing its increased tempertaure; we do not need to specify that the box is heated in a way such that the particles furthest from the center of the box increase their kinetic energy. This is because we assume that a hypothetical intervention heating a box of gas would also be an intervention bringing about a roughly uniform increase in the average kinetic energy of particles in all regions of the box, so that the scenario described above in which only the particles in the center of the box have increased energy tends not to occur. This kind of claim about what would happen under a hypothetical intervention increasing the temperature of a box of case is true just in case it can be derived from a Bayes net that truly represents the thermodynamics of the box of gas and includes both temperature and pressure as variables. Nothing about this is inconsistent with the fact that there may be some ways in which such a hypothetical intervention might be realized such that increasing the temperature of the box gas would not increase the pressure on the box.

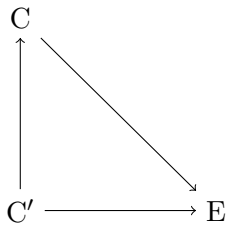

Figure 4: Multi-Level Causal Graph

I take these examples and results to motivate causal ecumenism as a philosophically and scientifically respectable thesis that is in keeping with both the Bayes nets formalism and the interventionist approach to causation. In what follows, I consider and respond to the primary argument in favor of the rival thesis that, according to the account of causation that one finds in the theory of Bayes nets, there is a uniquely correct level at which the causes of a given event can be described.

4 Against Causal Exclusion

Gebharter's (2017) argument for exclusion takes its cues from Kim, but uses the formal machinery of Bayes nets. His argument begins with the following stipulations: let E be some effect variable, and let C and C' be two possible causal variables such that C is a coarsening of C'. Gebharter posits that we can represent all three variables as part of the same graph, as shown in Figure 4. This stands in contrast with my account of coarse-graining, which represents different granularities of causation using different graphs. Gebharter points out that the graph in Figure 4 violates Minimality. To see why, note that even though there is an edge from C to E, the value of E is independent of the value of C, provided that we conditionalize on C'. This is because the probability distribution over both C and E will be totally determined by the value of C'. We can remove the edge from C to E, creating a proper sub-graph of the graph in Figure 4 that does not violate CMC. Thus, the graph in Figure 4 violates Minimality. By contrast, the proper subgraph induced by removing the edge from C to E does satisfy Minimality. Via this reasoning, Gebharter concludes that the causal relationship between C' and E excludes the possibility of a causal relationship between C and E, according the the Bayes nets axioms.

To illustrate the upshot of Gebharter's conclusion, suppose that we throw a ball at a window and it breaks. Suppose further that we know the microphysical properties of the ball as it makes impact

with the window. These microphysical properties allow us to assign a probability distribution over whether the window breaks. These microphysical properties also fix the macrophysical properties of the ball, such as its velocity and mass. So if we conditionalize on the microphysical properties of the ball, we see that whether or not the window breaks does not depend on the macrophysical properties of the ball. Thus, Gebharter would conclude, it cannot be the case that both the microphysical and macrophysical properties of the ball cause the window to break.

The premise that is doing the bulk of the work here is the claim that coarse-grained events and their fine-grained realizers should be represented using the same graph, and that fine-grained realizers should be represented as causing the coarse-grained properties that they realize. Gebharter's argument to this effect proceeds as follows. He begins by pointing out, correctly, that because Bayes nets satisfy CMC, the causal relations within a given Bayes Net have the property of *stability*. By stability, Gebharter means that for some causal relation $C \rightarrow E$, where C is the only parent of E, the conditional probability distribution over E, given C, is independent of the prior probability distribution over C. For example, the probability that a ball is thrown at a window, given that the window breaks, does not depend on the prior probability that the window breaks.[5] Gebharter argues that the supervenience relations that hold between coarse-grained events and their fine-grained realizers also exhibit this stability property. For example, the conditional probability that a ball has some velocity, given its microphysical properties, is not affected by the prior probability that the ball has its particular microphysical properties. Since this stability property seems to be an essential feature of both supervenience relations and causal relations, Gebharter concludes that supervenience relations and causal relations should have the same formal representation in a Bayesian network.

Up until his conclusion, everything that Gebharter says is correct. However, for his argument to be valid, it would have to be the case that sharing this stability property implies that two relations are indistinct. I argue that this assumption is false. Even though causation and supervenience relations both exhibit this stability property, there are other properties that distinguish the two types of relations. First, causal relations necessarily exhibit an asymmetry of probabilistic dependence under interventions. For example, intervening on the velocity of the ball thrown at a window changes the probability that the window breaks, but intervening so that a window does not break

[5]Note that this is distinct from the notion of "stability" in causal models discussed in Woodward (2010).

does not change the velocity of the ball thrown at it. This same asymmetry is not necessarily a property of supervenience relations; an intervention that changes the microphysical properties of the ball can also change its velocity, and a change in the velocity of the ball must result in a change in its microphysical realization.

Second, there are metaphysical reasons for thinking that supervenience relations are fundamentally different from causal relations. To begin, token causal relations hold between *distinct* events, whereas token supervenience relations only hold between properties of objects or events that are in some sense not wholly distinct. When I say that the solidity of my desk supervenes on its molecular structure, I am describing a relation between two properties of the same token object. It would not make sense to say that the molecular structure of my desk *causes* it to be solid. Similarly, if I throw a brick at a window and the window breaks, it would be natural to say that my throwing the brick caused the window to break, but wrong to say that the window breaking supervenes on my throwing the brick. Also, there is a difference regarding the role that temporal ordering plays in characterizing the two relations; even if we do not want to rule out simultaneous or backwards causation, it is certainly the case that many causes precede their effects temporally, whereas subvening and supervening events necessarily occur simultaneously. Gebharter disagrees with this line of argument, asserting that one could "describe the temperature of a gas in a tank as the effect of the behavior of the gas particles in the tank" (2017, p. 360). However, this assertion is presented without argument. While I agree that we might speak this way about temperature and gas particles, it is nevertheless strictly incorrect to do so, for the reasons given above. Although Gebharter is correct that supervenience and causation are both stable in the way that he describes, there remain important differences between the two relations (e.g. differences regarding symmetry or asymmetry of dependence under interventions on either relatum) that clearly matter in a Bayes nets context.

One might try to defend Gebharter here on the basis that supervenience and other constitutive relations are essentially "metaphysical" versions of causation, in contrast with "physical" causation, which holds between distinct events; a version of this view is defended by both Schaffer (2016) and Wilson (2018). My response here is just to say that even if supervenience is a metaphysical variant of causation, there is no reason to expect that a graph representing metaphysical causal relations satisfies CMC. Recall Pearl's justification of CMC: if all events are functions of their direct causes,

plus a variable-specific error term *that is not correlated with the error terms associated with other variables in the graph*, and all causes of two or more variables are included in the graph, then a probability distribution over the variables in the graph should satisfy CMC. However, in a graph containing variables that are metaphysically related to each other (e.g. a variable representing the velocity of each particle in a box of gas and a variable representing the temperature of the box of gas) these metaphysically related variables will have correlated error terms. For example, an exogenous source of indeterminacy that affects the velocity of each particle in the box of gas will also affect the temperature of the gas, since the former property of the gas ontologically depends on the latter. In this case, the model is what Pearl calls *semi-Markovian*, and it is not necessarily the case that the probability distribution over the variables in the graph satisfies CMC. Thus, we cannot assume CMC, and, therefore, Minimality, to hold in a Bayes net whose edges represent a distinctly metaphysical notion of causation, and so Gebharter's argument for causal exclusion using the Minimality condition does not go through. This same argument also explains why we cannot include a variable and its coarsening in the same graph when those two variables are related by a common cause, as opposed to being directly causally related to one another. Here again, we have reason to think that the error terms affecting the value of a variable and its coarsening will be correlated, such that there is no expectation that CMC will hold. Thus, the interventionist analysis of causation cannot be directly applied, and the causal significance of the graph, if any exists, is outside of the scope of the approach considered here.

Note that exogenous sources of error for a variable X' and its coarse-grained version X are correlated even in cases where values of X are multiply realized by values of X'. To see this, suppose that X' can take the values $\{1, 2, 3, 4\}$ and X can take the values $\{odd, even\}$. Clearly, not all changes in the value of X' change the value of X (e.g., if we change the value of X' from 1 to 3, the value of X will still be *odd*). Since X supervenes on X', there is a deterministic relationship between and X and X'. Nevertheless, exogenous sources of error in X' can be such that X' can be changed from an odd number to an even number, or vice-versa. Since such changes will necessarily lead to changes in the value of X, any error term for X will be correlated with the error in X' (indeed, they are effectively the same source of error). However, we might suppose that it is the case that the exogenous indeterminacy affecting X' is such that X' can only change from 1 to 3 or 3 to 1, or from 2 to 4 or 4 to 2. In such cases, there will be no exogenous indeterminacy in X

at all. However, this lack of exogenous indeterminacy in the value of X is *due to the nature of the indeterminacy in the value of* X', and so any error terms of the two variables are still correlated.

The correlation of error terms between a supervening variable and any variable that it supervenes upon holds even when the supervening variable coarsens the values of multiple fine-grained variables. To illustrate, let X be a coarse-grained variable, and let $\{X'_1, \ldots, X'_n\}$ be a set of fine-grained variables. Each fine-grained variable can take a value in the set $\{1, 2, \ldots, m\}$. However, if any fine-grained variable takes the value 1, it is necessarily the case that $X = 1$. As a real-world example of such a supervenience structure, it might be the case that the belief *Paris is the capital of France* is realized by wildly different physical states in a human, an extraterrestrial, an intelligent android, and even by significantly (though not wildly) different states in different humans, or in the same human at different times, and yet each of these states is itself sufficient for an intelligent being to have the thought *Paris is the capital of France*.[6] Let U'_i be the exogenous error term for each X'_i, and let U be the exogenous error term for X. Let us suppose that each variable can be represented as a function of only its error term. The suppositions made in this example imply that $p(U = 1|U'_i = 1) = 1$, and so the error terms are still correlated. This correlation exists in spite of the fact that any U'_i is uncorrelated with any U'_j such that $j \neq i$. To see this, note that $p(U = 1|U'_i = 1)$ can be re-written as follows:

$$p(U = 1|U'_i = 1) = \sum_{k=1}^{m} p(U = 1, U'_j = k|U'_i = 1) \text{ for any } j \neq i. \tag{7}$$

If for any $j \neq i$, U'_i is uncorrelated with U'_j, then we derive the following:

$$\sum_{k=1}^{m} p(U = 1, U'_j = k|U'_i = 1) = \sum_{k=1}^{m} p(U = 1|U'_i = 1)p(U'_j = k) = p(U = 1|U'_i = 1) = 1 \tag{8}$$

Thus, in such a scenario, each U'_i will be correlated with U, and so we cannot expect CMC to hold in general.

To illustrate how CMC can be violated in a Bayes net with uncorrelated error terms and deterministic relationships, consider a Bayes net $X \rightarrow Y \leftarrow Z$ in which Y is dependent (but not deterministically dependent) on X, Y is deterministically dependent on Z, and Y and Z have

[6] I am grateful to an anonymous reviewer for suggesting this example.

correlated error terms. Indeed, we suppose that once the value of Z is set via some exogenous, indeterministic process, the value of Y is set deterministically in virtue of Z taking its value, such that any exogenous source of indeterminism with respect to the value of Y is the very same source of indeterminism with respect to the value of Z. According to CMC, X and Z should be unconditionally independent of one another. However, under the suppositions just made, this independence claim does not hold. To see this, suppose that all three variables take values in the set $\{0, 1\}$. Suppose further that $p(X = 1) = .5$ (here the prior probability distribution over X is due solely to exogenous error), $p(X = 1|Y = 1) = .8$, and $p(Y = 1|Z = 1) = 1$. Under these conditions, $P(X = 1|Z = 1) = .8 \neq p(X = 1) = .5$, and so X and Z are not unconditionally independent. This result could be avoided if Y were to have an exogenous error term uncorrelated with the error in Z, such that $p(Y = 1|Z = 1) \neq 1$ and so fixing the value of Z would not also fix the value of Y. As discussed in Section 2, the standard justification of CMC assumes that all variables have such uncorrelated error terms, so that these sorts of scenarios are ruled out. However, Gebharter's use of the edges of a Bayes net to represent supervenience relations allows for precisely these sorts of CMC violations.

Gebharter might object here on the grounds that if $Z \rightarrow Y$ represents a supervenience relationship, then Y cannot have any parents other than the variable that it supervenes upon, since supervening variables do not stand in causal relationships to other variables. However, such a move presupposes exactly what causal exclusion arguments aim to establish (namely, that higher-level variables are causally inert), and is therefore question-begging. Gebharter might also object that the example seems to speak in favor of his thesis, rather than in opposition to it. What my example shows, he might argue, is that CMC fails in cases where supervening variables have parents other than the variables that they supervene upon, and thus the Bayes nets axioms speak in favor of causal exclusion. In response, I emphasize that *in general*, we cannot expect CMC to hold in contexts where variables do not have uncorrelated error terms, and so we still have no reason for thinking that CMC ought to hold in a Bayes net in which edges can represent either supervenience or causation.

My argument in this section is similar, in several respects, to another argument against Gebharter put forward by Stern and Eva (forthcoming). Namely, Stern and Eva agree with me that if, for a target given system, there is some representationally accurate Bayes net $\langle \mathcal{V}, \mathcal{E}, p(\cdot) \rangle$ such that

$X \in \mathcal{V}$ is a cause of $Y \in \mathcal{V}$, then X is a cause of Y. They call this the "Weak Causation Principle", and draw a contrast with the "Strong Causation Principle", which says that X is only a cause of Y if, for a given target system, in *every* representationally accurate Bayes net $\langle \mathcal{V}, \mathcal{E}, p(\cdot) \rangle$, $X \in \mathcal{V}$, $Y \in \mathcal{V}$, and X is a cause of Y.[7] They then show that even if we allow the edges of a Bayes net to represent relationships of both direct causation and supervenience, Gebharter's argument only goes through if we reject the Weak Causation Principle in favor of the Strong Causation Principle. They argue (and I agree) that such a rejection is unjustified. Indeed, rejecting the Strong Causation Principle in favor of the Weak Causation Principle is necessary in order for causal ecumenism to be a viable position, at least on my formalization of the thesis.

However, Stern and Eva also give a positive argument in favor of representing a given variable and a coarse-grained version of that variable in the same Bayes net. They suppose that such Bayes nets satisfy a version of CMC, in line with arguments put forward by Schaffer (2016) and Wilson (2018). However, they do not reckon with the fact that such Bayes nets will not generally be such that all error terms are uncorrelated. In light of my arguments above showing that such Bayes nets will not, in general, satisfy CMC, I take this supposition on their part to be unmotivated. Thus, while I agree with Stern and Eva's conclusion in favor of causal ecumenism, I disagree with some of their argumentative moves. Whereas they reject only the Strong Causation Principle, I reject both the Strong Causation Principle and the claim that the edges of a single Bayes net can be taken to represent either causation or supervenience.

5 The Nature of Macro-Level Interventions

One might object to the argument presented here on the grounds that, by defending causal ecumenism by citing interventional conditional probabilities whose conditioning events are interventions on macro-level variables, I am begging the question. The opponent of causal ecumenism, this line of argument goes, does not believe that there are genuine interventions bringing about macro-level events. All such interventions, the proponent of causal exclusion argues, are *really* interventions bringing about a micro-level, fine-grained event. Whether this is correct or not, what this line of reasoning putatively highlights is that what is at stake in the causal exclusion debate is

[7] Here I have translated Stern and Eva's thesis into my terminology.

precisely whether macro-level interventions are possible or well-defined. Thus, it could be argued, I beg the question by presupposing that such interventions are in fact well-defined.

In response, I argue that when we attempt to assess the truth or falsehood of a causal claim according to a Bayes net, concerns about the ontological status of an intervention are a red herring. An interventional conditional probability $p(y_j|do(x_i))$ is a mathematical object whose value can be derived from the observational joint probability distribution over the variables in a Bayes net that satisfies CMC. When $p(y_j|do(x_i)) > p(y_j)$ and the other conditions in Woodward's definition of actual causation are satisfied, the fact that the true probability distribution over the variables in the Bayes net satisfies CMC allows us to infer the true causal claim '$X = x_i$ causes $Y = y_j$.' This truth-preserving inference does not require us to consider the ontological status of the the intervention. Indeed, one of the advantages of the interventionist/Bayes nets approach to causation is that it allows us to assign values to interventional conditional probabilities even in cases where the intervention might be physically impossible to perform, (e.g. the interventions on atmospheric pressure calculated in Section 3). Thus, the result that Bayes nets support an ecumenist account of causation at multiple levels of description is not threatened by concerns about whether macro-level interventions can be ontologically well-defined; what matters is that they are mathematically well-defined. Proposition 2 shows that there are mathematically well-defined interventions, stated at different levels of granularity, that support an ecumenist approach to causation.

In this respect, I am actually in agreement with Gebharter, who claims, as I do, that "causation is characterized by means of the causal Markov condition and the causal minimality condition" (2017, p. 364). Like me, Gebharter is able to dispatch worries about the ontological status of interventions by treating CMC and Minimality as the ontologically motivated components of the Bayes nets approach to causation (i.e., both Gebharter and I assume that the world is such that an accurate representation of its causal structure must obey CMC and Minimality), and treating the interventionist semantics of causation as a useful corollary of CMC. As evidence that Gebharter and I share the view that one can be an instrumentalist (or at least be non-committal) about interventions while taking CMC and Minimality to be ontologically-motivated constraints on the representation of causal structure, note that Gebharter says the following, when discussing his argumentative approach in defending causal exclusion:

> In my argumentation I will not make use of intuitions about how interventions should work together with supervenience relationships. The reason for this is that there is still no consensus about that. It is still unclear whether directly intervening on mental properties is possible, or whether mental properties can only be manipulated by intervening on their physical supervenience bases, whether interventions on mental properties have to be common causes of these mental properties and their supervenience bases, etc. (2017, p. 358-9).

While Gebharter and I agree that debates about the ontological status of interventions are beside the point in the debate over causal exclusion, it should be clear from the above that I disagree with Gebharter with respect to whether supervenience and causation can both be represented using the graphical structure of a Bayes net. As argued above, there is no reason to expect CMC to hold when the arrows in a graph represent supervenience relations, because the error terms in variables related via supervenience will be correlated. Thus, if Gebharter is serious about causation being characterized by CMC, then he treats supervenience as a variety of causation on pain of inconsistency.

6 Conclusion

I have argued that the Bayes Nets approach to causal representation is generally consistent with an ecumenical approach to the relationship between granularity and causation. That is, there are many cases in which the causes of the same event can be described with varying degrees of granularity, such that both descriptions are correct. While this ecumenism does not hold in all cases, it holds in cases where the coarse-graining of the causal variable is done judiciously, so as to preserve the truth of certain causal claims. This ecumenical view stands in contrast with Gebharter's argument for causal exclusion in a Bayes nets context. I have argued that Gebharter's view faces serious issues, such that ecumenism is preferable.

7 Appendix

7.1 Calculations

7.1.1 Calculation of Inequality (4)

For the purpose of these calculations, *yes*, *no*, *low*, *medium*, and *high* are shortened to y, n, l, m and h, respectively. We begin with left-hand term of the inequality in (4). Since $S' \notin \{A'\}$ and A' is the sole parent of S', (2) and (3) jointly imply:

$$p(S' = y|do(A' = l)) = p(S' = y|A' = l) \tag{9}$$

Which can be calculated as follows:

$$p(S' = y|do(A' = l)) = \frac{p(A' = l, B' = h, S' = y) + p(A' = l, B' = l, S' = y)}{p(A' = l)} \tag{10}$$

$$p(S' = y|do(A' = l)) = \frac{.042 + .267}{p(A' = l)} \tag{11}$$

We can calculate $p(A' = l)$ as follows:

$$\begin{aligned} p(A' = l) = {} & p(A' = l, B' = h, S' = y) + p(A' = l, B' = h, S' = n) \\ & + \ p(A' = l, B' = l, S' = y) + p(A' = l, B' = l, S' = n) \end{aligned} \tag{12}$$

$$p(A' = l) = .042 + .003 + .267 + .021 = .333 \tag{13}$$

This yields the result:

$$p(S' = y|do(A' = l)) = \frac{.042 + .267}{.333} \approx .93 \tag{14}$$

Next, we calculate the right-hand term:

$$\begin{aligned} p(S' = y) = {} & p(A' = h, B' = h, S' = y) + p(A' = h, B' = l, S' = y) \\ & + p(A' = m, B' = h, S' = y) + p(A' = m, B' = l, S' = y) \\ & + p(A' = l, B' = h, S' = y) + p(A' = l, B' = l, S' = y) \end{aligned} \tag{15}$$

$$p(S' = y) = .042 + .003 + .1 + .067 + .042 + .267 = .52 \tag{16}$$

7.1.2 Calculation of Equation (5)

As we have already calculated the right-hand term, we focus on the left-hand term. From the law of total probability, we have:

$$\begin{aligned} p(S' = y|do(B' = l)) = p(S' = y, A' = l|do(B' = l)) \\ + \ p(S' = y, A' = m|do(B' = l)) + p(S' = y, A' = h|do(B' = l)) \end{aligned} \tag{17}$$

Since $S' \notin \{B'\}$, $A' \notin \{B'\}$, and A' is a parent of B', the equation above together with (2) and (3) jointly imply:

$$\begin{aligned} p(S' = y|do(B' = l)) = p(S' = y|A' = l)p(A' = l) \\ + \ p(S' = y|A' = m)p(A' = m) + p(S' = y|A' = h)p(A' = h) \end{aligned} \tag{18}$$

$$p(S' = y|do(B' = l)) = p(A' = l, S' = y) + p(A' = m, S' = y) + p(A' = h, S' = y) \tag{19}$$

To see this implication clearly, note that since $S' \notin \{B'\}$ and $A' \notin \{B'\}$, equation (2) and (3) imply that $p(S' = y, A' = l|do(B' = l)) = p(S' = y|\mathbf{pa}_{S'})p(A' = l|\mathbf{pa}_{A'})$. Since A' is the sole parent of S' and A' has no parents, $p(S' = y|\mathbf{pa}_{S'})p(A' = l|\mathbf{pa}_{A'}) = p(S' = y|A' = l)p(A' = l)$. We repeat these steps to obtain the other summands of (18). By the law of total probability, we obtain:

$$p(S' = y|do(B' = l)) = p(S' = y) \approx .52 \tag{20}$$

7.1.3 Calculation of Inequality (6)

We begin with left-hand term of the inequality in (6). Since $S' \notin \{A\}$ and A is the sole parent of S', (2) and (3) jointly imply:

$$p(S' = y|do(A = l/m)) = p(S' = y|A = l/m) \tag{21}$$

Which can be calculated as follows:

$$p(S' = y|do(A = l/m)) = \frac{p(A = l/m, B' = h, S' = y) + p(A = l/m, B' = l, S' = y)}{p(A = l/m)} \tag{22}$$

$$p(S' = y|do(A = l/m)) = \frac{.142 + .334}{p(A = l/m)} \tag{23}$$

We can calculate $p(A' = l/m)$ as follows:

$$\begin{aligned} p(A = l/m) = p(A = l/m, B' = h, S' = y) + p(A = l/m, B' = h, S' = n) \\ + \ p(A = l/m, B' = l, S' = y) + p(A = l/m, B' = l, S' = n) \end{aligned} \tag{24}$$

$$p(A = l/m) = .142 + .103 + .334 + .088 = .666 \tag{25}$$

This yields the result:

$$p(S' = y|do(A = l/m)) = \frac{.142 + .334}{.666} = .714 \tag{26}$$

Finally, we recall the value of the right-hand term $p(S' = y) = .52$.

7.2 Proof of Proposition 1

Proof. Assume $p(y_j|do(\mathbf{x}_i)) > p(y_j)$, where $\mathbf{x}_i$ is a vector of values that contains both x_i and the actual values taken by all parents of Y not on some directed path from X to Y. Let φ be the set of values of Y other than y_j. This implies that $1 - p(y_j|do(\mathbf{x}_i)) < 1 - p(\varphi)$, which implies in turn that there is a value y_l such that $p(y_l|do(\mathbf{x}_i)) < p(y_l)$. Suppose that the set $\mathcal{PA}_X$ containing all variables that are parents of X and all parents of Y not on the stipulated path from X to Y has the set of possible vectors of values $\{\mathbf{pa_{X1}}, \mathbf{pa_{X2}}, ..., \mathbf{pa_{Xq}}\}$. It is well known (see Pearl et al. 2016, p. 59) that we can derive the following:

$$p(y_l|do(\mathbf{x}_i)) = \sum_{t=1}^{q} p(y_l|\mathbf{x}_i, \mathbf{pa_{Xt}})p(\mathbf{pa_{Xt}}) < p(y_l) \tag{27}$$

The law of total probability implies the following, where X has n values:

$$p(y_l) = \sum_{k=1}^{n} \sum_{t=1}^{q} p(y_l|\mathbf{x}_i, \mathbf{pa_{Xt}}) p(\mathbf{x}_i, \mathbf{pa_{Xt}}) \tag{28}$$

This implies that there exists a set of values $\mathbf{x}_k$ such that $\mathbf{x}_i \cup \mathbf{x}_k \setminus \mathbf{x}_i \cap \mathbf{x}_k = \{x_i, x_k\}$, and:

$$\sum_{t=1}^{q} p(y_l|\mathbf{x}_i, \mathbf{pa_{Xt}}) p(\mathbf{pa_{Xt}}) < p(y_l) < \sum_{t=1}^{q} p(y_l|\mathbf{x}_k, \mathbf{pa_{Xt}}) p(\mathbf{x}_k, \mathbf{pa_{Xt}}) \tag{29}$$

The fact that $p(\mathbf{x}_i) > 0$ implies that $p(\mathbf{x}_k) < 1$, which implies in turn that $p(\mathbf{x}_k, \mathbf{pa_{Xt}}) < p(\mathbf{pa_{Xt}})$ for all $\mathbf{pa_{Xt}}$ with positive probability. It follows from this that:

$$p(y_l) < \sum_{t=1}^{q} p(y_l|\mathbf{x}_k, \mathbf{pa_{Xt}}) p(\mathbf{x}_k, \mathbf{pa_{Xt}}) < \sum_{t=1}^{q} p(y_l|\mathbf{x}_k, \mathbf{pa_{Xt}}) p(\mathbf{pa_{Xt}}) = p(y_l|do(\mathbf{x}_k)) \tag{30}$$

Which immediately implies that $p(y_l|do(x_k)) > p(y_l)$ when we hold fixed all parents of Y not on the stipulated path from X to Y. □

7.3 Proof of Proposition 2

Proof. Consider a Bayes net $\mathcal{N}' = \langle \mathcal{V}', \mathcal{E}', p(\cdot) \rangle$, with $C' \in \mathcal{V}'$, where $C' = c'_l$ causes $E' = e'_s$. This fact implies that $p(e'_s|do(c'_l, \mathbf{z})) > p(e'_s)$, where $\mathbf{z}$ is the actual setting of values for all parents of E' not on some directed path from C' to E'. Define an equivalence relation $\sim$ over the range of C' such that $c'_l \sim c'_u$ if and only if either:

a) $p(e'_s|do(c'_l, \mathbf{z})) > p(e'_s)$ and $p(e'_s|do(c'_u, \mathbf{z})) > p(e'_s)$, or

b) $p(e'_s|do(c'_l, \mathbf{z})) \leq p(e'_s)$ and $p(e'_s|do(c'_u, \mathbf{z})) \leq p(e'_s)$.

Let the range of C be a quotient set of the range of C' according to this equivalence relation. Define a Bayes net $\mathcal{N} = \langle \mathcal{V}, \mathcal{E}, p(\cdot) \rangle$ such that $\mathcal{V} = \{C\} \cup \mathcal{V}' \setminus \{C'\}$ and such that, for any edge $\langle V', C' \rangle \in \mathcal{E}'$ or $\langle C', V' \rangle \in \mathcal{E}'$, there is an edge $\langle V', C \rangle \in \mathcal{E}$ or $\langle C, V' \rangle \in \mathcal{E}$; all other elements of $\mathcal{E}'$ are included in $\mathcal{E}$. $\mathcal{N}$ trivially satisfies (1), (2) and (4); we now show that it satisfies (3) and (5) as well.

Let us begin with condition (3). To show that $\mathcal{N}$ satisfies CMC, let $X' \in \mathcal{V}'$ and $Y' \in \mathcal{V}'$ be two variables such that Y' is not a descendant of X' in $\mathcal{N}'$. If $X' \neq C'$, $Y' \neq C'$, and $C' \notin \mathcal{PA}_{X'}$,

then the supposition that $\mathcal{N}'$ satisfies CMC, along with the truth of conditions (1) and (2), implies that X' and Y' are independent, given $\mathcal{PA}_{X'}$, in $\mathcal{N}$.

If $C' = X'$, then the fact that $\mathcal{N}'$ satisfies CMC implies that for any values c'_l, y'_o, and vector of values $\mathbf{pa}_{\mathbf{C}'}$ of the variables in $\mathcal{PA}_{C'}$, $p(c'_l|y'_o, \mathbf{pa}_{\mathbf{C}'}) = p(c'_l|\mathbf{pa}_{\mathbf{C}'})$. If C is a coarsening of C', then for each value c_j, each conditional probability $p(c_j|y'_o, \mathbf{pa}_{\mathbf{C}})$ and $p(c_j|\mathbf{pa}_{\mathbf{C}})$ is a sum of terms of the form $p(c'_l|y'_o, \mathbf{pa}_{\mathbf{C}'})$ and $p(c'_l|\mathbf{pa}_{\mathbf{C}'})$, respectively. Thus, if the latter pair of terms are equal for all triples $(c'_l, y'_o, \mathbf{pa}_{\mathbf{C}'})$, then the former pair of terms are equal for all values $(c_j, y'_o, \mathbf{pa}_{\mathbf{C}})$. Thus, C is independent of its non-descendants, given its parents, in $\mathcal{N}$.

If $C' = Y'$, then the fact that $\mathcal{N}'$ satisfies CMC implies that for any values c'_l, x'_o, and set of values $\mathbf{pa}_{\mathbf{X}'}$ of the variables in $\mathcal{PA}_{X'}$, $p(x'_o|c'_l, \mathbf{pa}_{\mathbf{X}'}) = p(x'_o|\mathbf{pa}_{\mathbf{X}'})$. These conditional probabilities can be expressed as the following ratios:

$$p(x'_o|c'_l, \mathbf{pa}_{\mathbf{X}'}) = \frac{p(x'_o, c'_l, \mathbf{pa}_{\mathbf{X}'})}{p(c'_l, \mathbf{pa}_{\mathbf{X}'})} \tag{31}$$

$$p(x'_o|\mathbf{pa}_{\mathbf{X}'}) = \frac{p(x'_o, \mathbf{pa}_{\mathbf{X}})}{p(\mathbf{pa}_{\mathbf{X}'})} \tag{32}$$

If C is a coarsening of C', then for each value c_j, each joint probability $p(x'_o, c_j, \mathbf{pa}_{\mathbf{X}'})$ and $p(c_j, \mathbf{pa}_{\mathbf{X}'})$ is equal to a sum of joint probabilities of the form $p(x'_o, c'_l, \mathbf{pa}_{\mathbf{X}'})$ and $p(c'_l, \mathbf{pa}_{\mathbf{X}'})$, respectively. Thus, if $p(x'_o|c'_l, \mathbf{pa}_{\mathbf{X}'}) = p(x'_o|\mathbf{pa}_{\mathbf{X}'})$ for all triples $(c'_l, x'_o, \mathbf{pa}_{\mathbf{X}'})$, then $p(x'_o|c_j, \mathbf{pa}_{\mathbf{X}'}) = p(x'_o|\mathbf{pa}_{\mathbf{X}'})$ for all triples $(c_j, x'_o, \mathbf{pa}_{\mathbf{X}'})$. Thus, if C is a non-descendant of X' in $\mathcal{N}$, then X' is independent of C, given X''s parents, in $\mathcal{N}$. The immediately preceding analysis could be repeated if $C \in \mathcal{PA}_{X'}$, to show that any variable X' is independent of its non-descendants in $\mathcal{N}$, given its parents, when those parents include C. Together, these results show that $\mathcal{N}$ satisfies CMC. Minimality can be achieved by stipulation, by simply removing any edges that are not necessary for $\mathcal{N}$ to satisfy CMC.

Finally, we can show that (5) is true. Suppose that $C' = c'_l$ implies that $C = c_j$. In other words, c'_l is mapped to c_j in the coarsening function from the range of C' to the range of C. We know that $p(e'_s|do(c'_l, \mathbf{z})) > p(e'_s)$, and c'_l is $\sim$-related to all and only those values of C' such that conditioning on an intervention bringing about those values increases the probability that $E' = e'_s$, relative to its marginal probability. Thus, the conditional probability $p(e'_s|do(c_j, \mathbf{z}))$ is a sum of terms of the form $p(e'_s|do(c'_l, \mathbf{z}))$, each of which is such that $p(e'_s|do(c'_l, \mathbf{z})) > p(e'_s)$. This implies

that $p(e'_s|do(c_j, \mathbf{z})) > p(e'_s)$, and therefore that $C = c_j$ causes $E' = e'_s$. □

References

D. Davidson. Mental events. In L. Foster and J. W. Swanson, editors, *Experience and Theory*, pages 79–101. Humanities Press, 1970.

F. Eberhardt. Green and grue causal variables. *Synthese*, 193(4):1029–1046, 2016.

M. Eronen. Pluralistic physicalism and the causal exclusion argument. *European Journal for Philosophy of Science*, 2(2):219–232, 2012.

L. Fenton-Glynn. A proposed probabilistic extension of the halpern and pearl definition of 'actual cause'. *British Journal for the Philosophy of Science*, 68(4):1061–1124, 2017.

A. Gebharter. Causal exclusion and causal bayes nets. *Philosophy and Phenomenological Research*, 95(2):353–375, 2017.

C. Glymour. Review of woodward (2003). *British Journal for the Philosophy of Science*, 55: 779–790, 2004.

C. Glymour, D. Danks, B. Glymour, F. Eberhardt, J. Ramsey, R. Scheines, P. Spirtes, C. M. Teng, and J. Zhang. Actual causation: a stone soup essay. *Synthese*, 175(2):169–192, 2010.

D. Hausman and J. Woodward. Independence, invariance and the causal markov condition. *British Journal for the Philosophy of Science*, 50(4):521–583, 1999.

C. Hitchcock. Theories of causation and the causal exclusion argument. *Journal of Consciousness Studies*, 19(5-6):5–6, 2012.

V. Hoffmann-Kolss. Interventionism and higher-level causation. *International Studies in the Philosophy of Science*, 28(1):49–64, 2014.

Y. Huang and M. Valtorta. Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pages 217–224, 2006.

F. Jackson and P. Pettit. Functionalism and broad content. *Mind*, 97(July):318–400, 1988.

F. Jackson and P. Pettit. Causation and the philosophy of mind. *Philosophy and Phenomenological Research*, 50:195–214, 1990a.

F. Jackson and P. Pettit. Program explanation: A general perspective. *Analysis*, 50(2):107–17, 1990b.

F. Jackson and P. Pettit. In defense of explanatory ecumenism. *Economics and Philosophy*, 8(1): 1–21, 1992.

D. Janzing and B. Scholkopf. Causal inference using the algorithmic markov condition. *IEEE Transactions on Information Theory*, 56(10):5168–5194, 2010.

J. Kim. Mechanism, purpose, and explanatory exclusion. *Philosophical Perspectives*, 3:77–108, 1989.

J. Kim. *Mind in a physical world: An essay on the mind-body problem and mental causation.* MIT press, 2000.

C. List and P. Menzies. Nonreductive physicalism and the limits of the exclusion principle. *Journal of Philosophy*, 106(9):475–502, 2009.

J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, 2000.

J. Pearl, M. Glymour, and N. P. Jewell. *Causal inference in statistics: A primer.* John Wiley & Sons, 2016.

P. Pettit. The program model, difference-makers, and the exclusion problem. In H. P. H. Beebe, C. Hitchcock, editor, *Making a difference.* Oxford University Press, 2017.

T. W. Polger, L. A. Shapiro, and R. Stern. In defense of interventionist solutions to exclusion. *Studies in History and Philosophy of Science Part A*, 68:51–57, 2018.

P. Raatikainen. Causation, exclusion, and the special sciences. *Erkenntnis*, 73(3):349–363, 2010.

J. Schaffer. Grounding in the image of causation. *Philosophical studies*, 173(1):49–100, 2016.

S. Shimizu, P. O. Hoyer, A. Hyvärinen, and A. Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(Oct):2003–2030, 2006.

P. Spirtes. Variable definition and causal inference. *Proceedings of the International Congress for Logic, Methodology and Philosophy of Science*, 2007.

P. Spirtes and R. Scheines. Causal inference of ambiguous manipulations. *Philosophy of Science*, 71(5):833–845, 2004.

P. Spirtes, C. Glymour, S. N., and Richard. *Causation, Prediction, and Search*. MIT Press: Cambridge, 2000.

R. Stern and B. Eva. Antireductionist interventionism. *British Journal for the Philosophy of Science*, forthcoming.

B. Weslake. Explanatory depth. *Philosophy of Science*, 77(2):273–294, 2010.

A. Wilson. Metaphysical causation. *Noûs*, 52(4):723–751, 2018.

J. Woodward. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, 2003.

J. Woodward. Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology and Philosophy*, 25(3):287–318, 2010.

J. Woodward. Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, 91(2):303–347, 2015.

J. Zhang. A comparison of three occam's razors for markovian causal models. *The British Journal for the Philosophy of Science*, 64(2):423–448, 2012.

L. Zhong. Sophisticated exclusion and sophisticated causation. *The Journal of Philosophy*, 111(7): 341–360, 2014.